WO 02/05089 PCT/AU01/00799

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/05089 A1

(51) International Patent Classification[7]: G06F 9/45, 15/16, 17/00, 17/30, 17/60

(21) International Application Number: PCT/AU01/00799

(22) International Filing Date: 5 July 2001 (05.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PQ 8677 10 July 2000 (10.07.2000) AU

(71) Applicant *(for all designated States except US)*: CANON KABUSHIKI KAISHA [JP/JP]; 30-2, Shimomaruko 3-chome, Ohta-ku, Tokyo 146 (JP).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: WAN, Ernest, Yiu, Cheong [AU/AU]; 2A Wilshire Avenue, Carlingford, NSW 2118 (AU).

(74) Agent: SPRUSON & FERGUSON; GPO Box 3898, Sydney, NSW 2001 (AU).

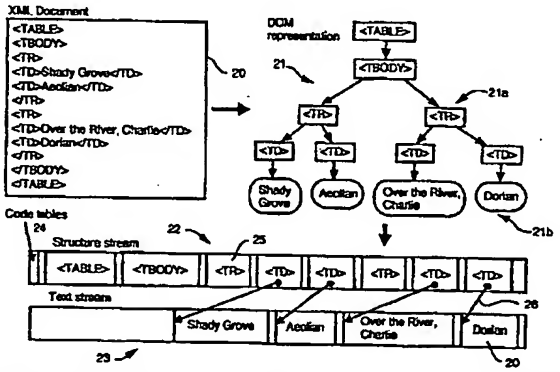
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DELIVERING MULTIMEDIA DESCRIPTIONS



(57) Abstract: Disclosed is method of processing a document (20) described in a mark up language (eg. XML). Initially, a structure (21a) and a text content (21b) of the document are separated, and then the structure (22) is transmitted, for example by streaming, before the text content (23). Parsing of the received structure (22) is commenced before the text content (23) is received. Also disclosed is a method of forming a streamed presentation (37, 38) from at least one media object having content (31, 32) and description (33) components. A presentation description (35) is generated (36) from at least one component description of the media object and is then processed (34) to schedule delivery of component descriptions and content of the presentation to generate elementary data streams associated with the component descriptions (38) and content (37). Another method of forming a streamed presentation of at least one media object having content and description components is also disclosed. A presentation template (53) is provided that defines a structure of a presentation description (56). The template is then applied (54) to at least one description component (52) of the associated media object to form the presentation description from each description component. The presentation description is then stream encoded with each associated media object (51) to form the streamed presentation (57, 58), whereby the media object is reproducible using the presentation description.

WO 02/05089 A1

DELIVERING MULTIMEDIA DESCRIPTIONS

Technical Field of the Invention

The present invention relates generally to the distribution of multimedia and, in particular, to the delivery of multimedia descriptions in different types of applications. The present invention has particular application to, but is not limited to, the evolving MPEG-7 standard.

Background Art

Multimedia may be defined as the provision of, or access to, media, such as text, audio and images, in which an application can handle or manipulate a range of media types. Invariably where access to a video is desired, the application must handle both audio and images. Often such media is accompanied by text that describes the content and may include references to other content. As such, multimedia may be conveniently referred to as being formed of content and descriptions. The description is typically formed by metadata which is, practically speaking, data which is used to described other data.

The World Wide Web (WWW or, the "Web") uses a client/server paradigm. Traditional access to multimedia over the Web involves an individual client accessing a database available via a server. The client downloads the multimedia (content and description) to the local processing system where the multimedia may be utilised, typically by compiling and replaying the content with the aid of the description. The description is "static" in that usually the entire description must be available at the client in order for the content, or parts thereof, to be reproduced. Such traditional access is problematic in the delay between client request and actual reproduction, and the sporadic load on both the server and any communications network linking the server and local

processing system as media components are delivered. Real-time delivery and reproduction of multimedia in this fashion is typically unobtainable.

The evolving MPEG-7 standard has identified a number of potential applications for MPEG-7 descriptions. The various MPEG-7 "pull", or retrieval applications, involve client access to databases and audio-visual archives. The "push" applications are related to content selection and filtering and are used in broadcasting, and the emerging concept of "webcasting", in which media, traditionally broadcast over the airways by radio frequency propagation, is broadcast over the structured links of the Web. Webcasting, in its most fundamental form, requires a static description and streamed content. However webcasting usually necessitates the downloading of the entire description before any content may be received. Desirably, webcasting requires streamed descriptions received with or in association with, the content. Both types of applications benefit strongly from the use of metadata.

The Web is likely to be the primary medium for most people to search and retrieve audio-visual (AV) content. Typically, when locating information, the client issues a query and a search engine searches its database and/or other remote databases for relevant content. MPEG-7 descriptions, which are constructed using XML documents, enable more efficient and effective searching because of the well-known semantics of the standardised descriptors and description schemes used in MPEG-7. Nevertheless, MPEG-7 descriptions are expected to form only a (small) portion of all content descriptions available on the Web. It is desirable for MPEG-7 descriptions to be searchable and retrievable (or downloadable) in the same manner as other XML documents on the Web since users of the Web do not expect or want AV content to be downloaded with description. In some cases, the descriptions rather than the AV content

are what may be required. In other cases, users will want to examine the description before deciding on whether to download or stream the content.

MPEG-7 descriptors and description schemes are only a sub-set of the set of (well-known) vocabulary used on the Web. Using the terminology of XML, the MPEG-7 descriptors and description schemes are elements and types defined in the MPEG-7 namespace. Further, Web users would expect that MPEG-7 elements and types could be used in conjunction with those of other namespaces. Excluding other widely used vocabularies and restricting all MPEG-7 descriptions to consist only of the standardised MPEG-7 descriptors and description schemes and their derivatives would make the MPEG-7 standard excessively rigid and unusable. A widely accepted approach is for a description to include vocabularies from multiple namespaces and to permit applications to process elements (from any namespace, including MPEG-7) that the application understands, and ignore those elements that are not understood.

To make downloading, and any consequential storing, of a multimedia (eg. MPEG-7) description more efficient, the descriptions can be compressed. A number of encoding formats have been proposed for XML, and include WBXML, derived from the Wireless Application Protocol (WAP). In WBXML, frequently used XML tags, attributes and values are assigned a fixed set of codes from a global code space. Application specific tag names, attribute names and some attribute values that are repeated throughout document instances are assigned codes from some local code spaces. WBXML preserves the structure of XML documents. The content as well as attribute values that are not defined in the Document Type Definition (DTD) can be stored in line or in a string table. An example of encoding using WBXML is shown in Figs. 1A and 1B. Fig. 1A depicts how an XML source document 10 is processed by an interpreter 14 according various code spaces 12 defining encoding rules for WBXML. The interpreter 14 produces an

encoded document 16 suitable for communication according to the WBXML standard. Fig. 1B provides a description of each token in the data stream formed by the document 16.

While WBXML encodes XML tags and attributes into tokens, no compression is performed on any textual content of the XML description. Such may be achieved using a traditional text compression algorithm, preferably taking advantage of the schema and data-types of XML to enable better compression of attribute values that are of primitive data-types.

Summary of the Invention

It is an object of the present invention to substantially overcome, or at least ameliorate, one or more disadvantages of existing arrangements to support the streaming of multimedia descriptions.

General aspects of the present invention provide for streaming descriptions, and for streaming descriptions with AV (audio-visual) content. When streaming descriptions with AV content, the streaming can be "description-centric" or "media-centric". The streaming can also be unicast with upstream channel or broadcast.

According to a first aspect of the invention, there is provided a method of forming a streamed presentation from at least one media object having content and description components, said method comprising the steps of:

generating a presentation description from at least one component description of said at least one media object; and

processing said presentation description to schedule delivery of component descriptions and content of said presentation to generate elementary data streams associated with said component descriptions and content.

According to another aspect of the present invention there is disclosed a method of forming a presentation description for streaming content with description, said method comprising the steps of:

providing a presentation template that defines a structure of a presentation description;

applying said template to at least one description component of at least one associated media object to form said presentation description from each said description component, said presentation description defining a sequential relationship between description components desired for streamed reproduction and content components associated with said desired descriptions.

According to another aspect of the present invention there is disclosed a streamed presentation comprising a plurality of content objects interspersed amongst a plurality of description objects, said description objects comprising references to multimedia content reproducible from said content objects.

According to another aspect of the present invention there is disclosed a method of delivering an XML document, said method comprising the steps of:

dividing the document to separate XML structure from XML text; and

delivering said document in a plurality of data streams, at least one said stream comprising said XML structure and at least one other of said streams comprising said XML text.

In accordance with another aspect of the present invention, there is disclosed a method of processing a document described in a mark up language, said method comprising the steps of:

separating a structure and a text content of said document;

sending the structure before the text content; and

commencing to parse the received structure before the text content is received.

Other aspects of the present invention are also disclosed.

Brief Description of the Drawings

At least one embodiment of the present invention will now be described with reference to the drawings, in which:

Figs. 1A and 1B show an example of a prior art encoding of an XML document;

Fig. 2 illustrates a first method of streaming an XML document;

Fig. 3 illustrates a second method of "description-centric" streaming in which the streaming is driven by a presentation description;

Fig. 4A illustrates a prior art stream;

Fig. 4B shows a stream according to one implementation of the present disclosure;

Fig. 4C shows a preferred division of a description stream;

Fig. 5 illustrates a third method of "media-centric" streaming;

Fig. 6 is an example of a composer application;

Fig. 7 is a schematic block diagram of a general purpose computer upon which the implementation of the present disclosure can be practiced; and

Fig. 8 schematically represents an MPEG-4 stream.

Detailed Description including Best Mode

The implementations to be described are each founded upon the relevant multimedia descriptions being XML documents. XML documents are mostly stored and transmitted in their raw textual format. In some applications, XML documents are compressed using some traditional text compression algorithms for storage or transmission, and decompressed back into XML before they are parsed and processed. Although compression may greatly reduce the size of an XML document, and thus reduce the time for reading or transmitting the document, an application still has to receive the

entire XML document before the document can be parsed and processed. A traditional XML parser expects an XML document to be well-formed (ie. the document has matching and non-overlapping start-tag and end-tag pairs), and is unable to complete the parsing of the XML document until the whole XML document is received. Incremental parsing of a streamed XML document is unable to be performed using a traditional XML parser.

Streaming an XML document permits parsing and processing to commence as soon as a sufficient portion of the XML document is received. Such capability will be most useful in the case of a low bandwidth communication link and/or a device with very limited resources.

One way of achieving incremental parsing of an XML document is to send the tree hierarchy of an XML document (such as the Dominant Object Model (DOM) representation of the document) in a breadth-first or depth-first manner. To make such a process more efficient, the XML (tree) structure of the document can be separated from the text components of the document and encoded and sent before the text. The XML structure is critical in providing the context for interpreting the text. Separating the two components allows the decoder (parser) to parse the structure of the document more quickly, and to ignore elements that are not required or are unable to be interpreted. Such a decoder (parser) may optionally choose not to buffer any irrelevant text that arrives at a later stage. Whether the decoder converts the encoded document back into XML or not depends on the application.

The XML structure is vital in the interpretation of the text. In addition, as different encoding schemes are usually used for the structure and the text and, in general, there is far less structural information than textual content, two (or more) separate streams may be used for delivering the structure and the text.

Fig. 2 shows one method of streaming XML document 20. Firstly, the document 20 is converted to a DOM representation 21, which is then streamed in a depth-first fashion. The structure of the document 20, depicted by the tree 21a of the DOM representation 21, and the text content 21b, are encoded as two separate streams 22 and 23 respectively. The structure stream 23 is headed by code tables 24. Each encoded node 25, representing a node of the DOM representation 21, has a *size* field that indicates its size including the total size of corresponding descendant nodes. Where appropriate, encoded leaf nodes and attribute nodes contain pointers 26 to their corresponding encoded content 27 in the text stream 23. Each encoded string in the text stream is headed by a *size* field that indicates the size of the string.

Not all multimedia (eg. MPEG-7) descriptions need be streamed with content or serve as a presentation. For instance, television and film archives store a vast amounts of multimedia material in several different formats, including analogue tapes. It would not be possible to stream the description of a movie, in which the movie is recorded on analogue tapes, with the actual movie content. Similarly, treating the multimedia description of a patient's medical records as a multimedia presentation makes little sense. As an analogy, while Synchronised Multimedia Integration Language (SMIL) presentations are themselves XML documents, not all XML documents are SMIL presentations. Indeed, only a very small number of XML documents are SMIL presentations. SMIL can be used for creating presentation script that enables a local processor to compile an output presentation from a number of local files or resources. SMIL specifies the timing and synchronisation model but does not have any built-in support for the streaming of content or description.

Fig. 3 shows an arrangement 30 for streaming descriptions together with content. A number of multimedia resources are shown including audio files 31 and video files 32.

Associated with the resources 31 and 32 are descriptions 33 each typically formed of a number of descriptors and descriptor relationships. Significantly, there need not be a one-to-one relationship between the descriptions 33 and the content files 31 and 32. For example, a single description may relate to a number of files 31 and/or 32, or any one file 31 or 32 may have associated therewith more than one description.

As seen in Fig. 3, a presentation description 35 is provided to describe the temporal behaviour of a multimedia presentation desired to be reproduced through a method of description-centric streaming. The presentation description 35 can be created manually or interactively through the use of editing tools and a standardized presentation description scheme 36. The scheme 36 utilises elements and attributes to define the hyperlinks between the multimedia objects and the layout of the desired multimedia presentation. The presentation description 35 can be used to drive the streaming process. Preferably, the presentation description is an XML document that uses a SMIL-based description scheme.

An encoder 34, with knowledge of the presentation description scheme 36, interprets the presentation description 35, to construct an internal time graph of the desired multimedia presentation. The time graph forms a model of the presentation schedule and synchronization relationships between the various resources. Using the time graph, the encoder 34 schedules the delivery of the required components and then generates elementary data streams 37 and 38 that may be transmitted. Preferably, the encoder 34 splits the descriptions 33 of the content into multiple data streams 38. The encoder 34 preferably operates by constructing a URI table that maps the URI-references contained in the AV content 31, 32 and the descriptions 33 to a local address (eg. offset) in the corresponding elementary (bit) streams 37 and 38. The streams 37 and 38, having

been transmitted, are received into a decoder (not illustrated) that uses the URI table when attempting to decode any URI-reference.

The presentation description scheme 36, in some implementations, may be based on SMIL. Current developments in MPEG-4 enable SMIL-based presentation description to be processed into MPEG-4 streams.

An MPEG-4 presentation is made up of scenes. An MPEG-4 scene follows a hierarchical structure called a scene graph. Each node of the scene graph is a compound or primitive media object. Compound media objects group primitive media objects together. Primitive media objects correspond to leaves in the scene graph and are AV media objects. The scene graph is not necessarily static. Node attributes (eg. positioning parameters) can be changed and nodes can be added, replaced or removed. Hence, a scene description stream may be used for transmitting scene graphs, and updates to scene graphs.

An AV media object may rely on streaming data that is conveyed in one or more elementary streams (ES). All streams associated to one media object are identified by an object descriptor (OD). However, streams that represent different content must be referenced through distinct object descriptors. Additional auxiliary information can be attached to an object descriptor in a textual form as an OCI (object content information) descriptor. It is also possible to attach an OCI stream to the object descriptor. The OCI stream conveys a set of OCI events that are qualified by their start time and duration. The elementary streams of an MPEG-4 presentation are schematically illustrated in Fig. 8.

In MPEG-4, information about an AV object is stored and transmitted using the Object Content Information (OCI) descriptor or stream. The AV object contains a reference to the relevant OCI descriptor or stream. As seen in Fig. 4A, such an

arrangement requires a specific temporal relationship between the description and the content and a one-to-one relationship between AV objects and OCI.

However, typically, multimedia (eg. MPEG-7) descriptions are not written for specific MPEG-4 AV objects or scene graphs and, indeed are written without any specific knowledge of the MPEG-4 AV objects and scene graphs that make up the presentation. The descriptions usually provide a high level view of the information of the AV content. Hence, the temporal scope of the descriptions might not align with those of the MPEG-4 AV objects and scene graphs. For instance, a video/audio segment described by an MPEG-7 description may not correspond to any MPEG-4 video/audio stream or scene description stream. The segment may describe the last portion of one video stream and the beginning part of the following one.

The present disclosure presents a more flexible and consistent approach in which the multimedia description, or each fragment thereof, is treated as another class of AV object. That is, like other AV objects, each description will have its own temporal scope and object descriptor (OD). The scene graph is extended to support the new (eg. MPEG-7) description node. With such a configuration, it is possible to send a multimedia (eg. MPEG-7) description fragment, that has sub-fragments of different temporal scopes, as a single data stream or as separate streams, regardless of the temporal scopes of the other AV media objects. Such a task is performed by the encoder 34 and a example of such a structure, applied to the MPEG-4 example of Fig. 4A, is shown in Fig. 4B. In Fig. 4B, the OCI stream is also used to contain references of relevant description fragments and other AV object specific information as required.

Treating MPEG-7 descriptions in the same way as other AV objects also means that both can be mapped to a media object element of the presentation description scheme 36 and subjected to the same timing and synchronisation model. Specifically, in

the case of an SMIL-based presentation description scheme 36, a new media object element, such as an <mpeg7> tag, may be defined. Alternately, MPEG-7 descriptions can be treated as a specific type of text (eg. represented in Italics). Note that a set of common media object elements <video>, <audio>, <animation>, <text>, etc. are pre-defined in SMIL. The description stream can potentially be further separated into a structure stream and a text stream.

In Fig. 4C, a multimedia stream 40 is shown which includes an audio stream 41 and a video stream 42. Also included is a high-level scene description stream 46 comprising (compound or primitive) nodes of media objects and having leaf nodes (which are primitive media objects) that point to object descriptors ODn that make up an object descriptor stream 47. A number of low level description streams 43, 44 and 45 are also shown, each having components configured to be pointed to, or linked to the object description stream 47, as do the audio and video streams 41 and 42. With such an object-oriented streaming treating both content and description as media objects, the temporally irregular relationship between description and content may be accommodated through a temporal object description structured into the streams.

The above approach to streaming descriptions with content is appropriate where the description has some temporal relationship with the content. An example of this is a description of a particular scene in a movie, that provides for multiple camera angles to be viewed, thus permitting viewer access to multiple video streams for which only one video stream may, practically speaking, be viewed in the real-time running of the movie. This is to be contrasted with arbitrary descriptions which have no definable temporal relationship with the streamed content. An example of such may be a newspaper critic's text review of the movie. Such a review may make text reference, as opposed to a temporal and spatial reference to scenes and characters. Converting an arbitrary

description into a presentation is a non-trivial (and often impossible) task. Most descriptions of AV content are not written with presentation in mind. They simply describe the content and its relationship with other objects at various levels of granularity and from different perspectives. Generating a presentation from a description that does not use the presentation description scheme 36 involves arbitrary decisions, best made by a user operating a specific application, as opposed to the systematic generation of the presentation description 35.

Fig. 5 shows another arrangement 50 for streaming descriptions with content that the present inventor has termed "media-centric". AV content 51 and descriptions 52 of the content 51 are provided to a composer 54, also input with a presentation template 53 and having knowledge of a presentation description scheme 55. Although the content 51 shows a video and its audio track is shown as the initial AV media object, the initial AV object can actually be a multimedia presentation.

In media-centric streaming, an AV media object provides the AV content 51 and the timeline of the final presentation. This is in contrast to the description centric streaming where the presentation description provides the timeline of the presentation. Information relevant to the AV content is pulled in from a set of descriptions 52 of the content by the composer 54 and delivered with the content in a final presentation. The final presentation output from the composer 54 is in the form of elementary streams 57 and 58, as with the previous configuration of Fig. 3, or as a presentation description 56 of all the associated content.

The presentation template 53 is used to specify the type of descriptive elements that are required and those that should be omitted for the final presentation. The template 53 may also contain instructions as to how the required descriptions should be incorporated into the presentation. An existing language such as XSL Transformations (XSLT) may

be used for specifying the templates. The composer 54, which may be implemented as a software application, parses the set of required descriptions that describe the content, and extracts the required elements (and any associated sub-elements) to incorporate the elements into the time line of the presentation. Required elements are preferably those elements that contain descriptive information about the AV content that is useful for the presentation. In addition, elements (from the same set of the descriptions) that are referred to (by IDREF's or URI-references) by the selected elements are also included and streamed before their corresponding referring elements (their "referrers"). It is possible that a selected element is in turn referenced (either directly or indirectly) by an element that it references. It is also possible that a selected element has a forward reference to another selected element. An appropriate heuristic may be used to determine the order by which such elements are streamed. The presentation template 53 can also be configured to avoid such situations.

The composer 54 may generate the elementary streams 57, 58 directly, or output the final presentation as the presentation description 56 that conforms to the known presentation description scheme 55.

Fig. 6 is an example showing how the composer application 54 uses an XSLT-based presentation template 60 to extract the required description fragments from a movie description 62 to generate a SMIL-like presentation description 64 (or presentation script). The <par> container of SMIL specifies the start time and duration of a set of media objects that are to be presented in parallel. The <mpeg7> element shown in the presentation description 64 for example identifies the MPEG-7 description fragments. The description may be provided in-line or referred to by an URI reference. The *src* attribute contains an URI reference to the relevant description (fragment). The *content* attribute of the presentation description 64 describes the context of the included

description. Special elements, such as an <mpeg7> tag, can be defined in the presentation description scheme 55 for specifying description fragments that can be streamed separately and/or at different times in the presentation description 64.

The use of the presentation description schemes 36 and 55, each as a multimedia presentation authoring language, bridges the two described methods of description-centric and media-centric streaming. The schemes 36 and 55 also allow for a clear separation between the application and the system layer to be made. Specifically, the composer application 54 of Fig. 5, when outputting the presentation as a (presentation) description 56 permits the description 56 be used as the input presentation description 35 in the arrangement of Fig. 3, thereby permitting an encoder 34 residing at the system layer to generate the required elementary streams 37, 38 from the presentation description 56.

In the case of streaming description with AV content, it is questionable whether a very efficient means of compressing the description is required as the size of the description is likely to be insignificant when compared to that of the AV content. Nevertheless, streaming of the description is still necessary because transmitting (and, in case of broadcasting, repeating) the entire description before the AV content may result in high latency and require a large buffer at the decoder.

For a description that forms part of a multimedia presentation, it may appear that the corresponding content changes along the presentation's timeline. The description, however, is not really "dynamic" (ie. it does not change with time). More correctly, different information from different descriptions or different parts of a description are being delivered and incorporated into the presentation at different times. Actually, if enough resources and bandwidth are available, all the "static" descriptions could be sent to the receiver at the same time for incorporating into a presentation at a later time.

Nevertheless, the information delivered and presented during the presentation may be considered as generating a transient "dynamic" description.

If most of the information presented from one time instance to the next time instance remain unchanged, updates can be sent to effect the changes without repeating the unchanged information. The presented elements may be tagged with a begin time and a duration (or end time) just like other AV objects. Other attributes such as the position (or the context) of the element can also be specified. One possible approach is to use an extension of SMIL for specifying the timing and synchronization of the AV objects and the (fragments of) descriptions.

For example, the fragments of descriptions that go with a video clips of a soccer team may be specified according to Example 1 of SMIL-like XML code below:

Example 1:

```
<!-- Description of the team is relevant during the team's video clip -->
<par begin="teamAIntroductionVideo.begin" end="teamAIntroductionVideo.end">
  <text src="soccerTeam/teamA.xml#pointer(/soccerTeam/teamInfo)"
      context="/soccerTeam/teamInfo"/>
  <!-- Descriptions of the players are presented.
      Each last for 15 seconds. -->
  <seq>
    <text src="soccerTeam/teamA.xml#xpointer(/soccerTeam/player[1])"
        dur="15s" context="/soccerTeam/player"/>
    <text src="soccerTeam/teamA.xml#xpointer(/soccerTeam/player[2])"
        dur="15s" context="/soccerTeam/player"/>
    ...
  </seq>
</par>
```

Updates to a "dynamic" description have to be applied with care. A partial update might leave the description in an inconsistent state. For video and audio, packets of data

lost during transmission over the Web mostly appear as noise or even go unnoticed. However, inconsistent description may lead to wrong interpretations with serious consequences. For instance, in a weather report, if after the city element of a description is updated from "Tokyo" to "Sydney", the update to the temperature element was lost, the description would report the temperature of Tokyo as the temperature of Sydney. As another example, if after updating the coordinates of an approaching aircraft in a streamed video game, the category element of the description is lost, a "friendly" aircraft might be mistakenly labelled as "hostile".

As yet another example, shown in Example 2 below, an item number in a sale catalogue may become tagged with the wrong price. Hence, all related updates to a description have to be applied at once, or within a well-defined period, or not at all. For instance, in the following sales catalogue examples, every 10 seconds, the matching description and price of a new item is presented. The SMIL element *par* is used to hold all the related descriptive elements. A new *sync* attribute is used to make sure that matching description and price will be presented or not at all. The *dur* attribute makes sure that the information is applied for an appropriate period of time and then removed from the display.

Example 2:

```
<!--
    A sales catalogue. Each item on sale is presented for 10 seconds.
    More complex synchronization model can be specified, for instance,
    the begin and end time of each par container can be synchronized
    with that of a video clip of the item.
-->
<seq>
  <par dur="10s" sync="true">
    <text src="products.xml#xpointer(/products/item[1]/description)"
        context="/products/item/description"/>
```

```
        <text src="products.xml#xpointer(/products/item[1]/price)"
            context="/product/item/description"/>
      </par>
      <par dur="10s" sync="true">
        <text src="products.xml#xpointer(/products/item[2]/description)"
            context="/products/item/description"/>
        <text src="products.xml#xpointer(/products/item[2]/price)"
            context="/products/item/price"/>
      </par>
      ...
    </seq>
```

A streaming decoder has to buffer the *synced* set of elements and apply them as a whole. Missing information can be tolerated, as long as the incomplete information is consistent, and the *sync* attribute will not be required. In such cases, related elements can also be delivered and/or presented over a period of time. This can be demonstrated using Example 3 below:

Example 3:

```
    <!--
        A sales catalogue.  Each item on sale is presented for 10 seconds.
        The price is only made available 3 seconds after its description.
        (N.B. timing information relating to a set of updates is only
        useful if the elements are mapped directly to text on the screen.)
    -->
    <seq>
      <par dur="10s">
        <text src="products.xml#xpointer(/products/item[1]/description)"
            region="description"
            context="/products/item/description" />
        <text src="products.xml#xpointer(/products/item[1]/price)"
            region="price"
            context="/products/item/price"
            begin="3s" />
```

```
    </par>
    <par dur="10s">
      <text src="products.xml#xpointer(/products/item[2]/description)"
          region="description"
          context="/products/item/description"/>
      <text src="products.xml#xpointer(/products/item[2]/price)"
          region="price"
          context="/products/item/price"
          begin="3s" />
    </par>
    ...
  </seq>
```

It is extremely difficult, if not impossible, to decide at the system layer what updates to the document-tree are related and should be grouped without any hints from the description. Hence, while the system layer may allow updates to be grouped in the data streams and provide a means (such as the *sync* attribute in the above presentation description examples) to allow application to specify such grouping, the exact grouping should be left to the specific application.

If an upstream channel is available from the client to the server, the client can choose to signal the server for any lost or corrupted updated packets and request for their re-transmission, or ignore the entire set of updates.

In cases where the description is broadcast with AV content, the XML structure and text of the description should desirably be repeated at regular intervals throughout the duration that the description is relevant to the AV content. This allows the users to access (or tune into) the description at a time not predetermined. The description does not have to be repeated as frequently as the AV content because the description changes much less frequently and, at the same time, consumes significantly fewer computing resources at the decoder end. Nevertheless, the description should be repeated frequently enough so that

users are able to use the description without perceptible delay after tuning into the broadcast program. If the description changes at about the same rate at which it is repeated, or at a lower rate, then it is questionable that the ability to "dynamically" update the description is important or actually required.

The methods of streaming descriptions with content described above may be practiced using a general-purpose computer system 700, such as that shown in Fig. 7 wherein the processes of Figs. 2 to 6 may be implemented as software, such as an application program executing within the computer system 700. In particular, the steps of methods are effected by instructions in the software that are carried out by the computer. The software may be divided into two separate parts; one part for carrying out the encoding/composing/streaming methods; and another part to manage the user interface between the former and the user. The software may be stored in a computer readable medium, including the storage devices described below, for example. The software is loaded into the computer from the computer readable medium, and then executed by the computer. A computer readable medium having such software or computer program recorded on it is a computer program product. The use of the computer program product in the computer preferably effects an advantageous apparatus for description with content streaming in accordance with the embodiments of the invention.

The computer system 700 comprises a computer module 701, input devices such as a keyboard 702 and mouse 703, output devices including a printer 715 and a display device 714. A Modulator-Demodulator (Modem) transceiver device 716 is used by the computer module 701 for communicating to and from a communications network 720, for example connectable via a telephone line 721 or other functional medium. The modem 716 can be used to obtain access to the Internet, and other network systems, such as a Local Area Network (LAN) or a Wide Area Network (WAN). It is via the

device 716 that streamed multimedia may be broadcast or webcast from the computer module 701.

The computer module 701 typically includes at least one processor unit 705, a memory unit 706, for example formed from semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 707, and an I/O interface 713 for the keyboard 702 and mouse 703 and optionally a joystick (not illustrated), and an interface 708 for the modem 716. A storage device 709 is provided and typically includes a hard disk drive 710 and a floppy disk drive 711. A magnetic tape drive (not illustrated) may also be used. A CD-ROM drive 712 is typically provided as a non-volatile source of data. The components 705 to 713 of the computer module 701, typically communicate via an interconnected bus 704 and in a manner which results in a conventional mode of operation of the computer system 700 known to those in the relevant art. Examples of computer platforms on which the embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom, particularly when provided as a server incarnation.

Typically, the application program of the preferred embodiment is resident on the hard disk drive 710 and read and controlled in its execution by the processor 705. Intermediate storage of the program and any data fetched from the network 720 may be accomplished using the semiconductor memory 706, possibly in concert with the hard disk drive 710. The hard disk drive 710 and the CD-ROM 712 may form sources for the multimedia description and content information. In some instances, the application program may be supplied to the user encoded on a CD-ROM or floppy disk and read via the corresponding drive 712 or 711, or alternatively may be read by the user from the network 720 via the modem device 716. Still further, the software can also be loaded into

the computer system 700 from other computer readable medium including magnetic tape, a ROM or integrated circuit, a magneto-optical disk, a radio or infra-red transmission channel between the computer module 701 and another device, a computer readable card such as a PCMCIA card, and the Internet and Intranets including e-mail transmissions and information recorded on websites and the like. The foregoing is merely exemplary of relevant computer readable media. Other computer readable media may be practiced without departing from the scope and spirit of the invention.

Some aspects of the streaming methods may be implemented in dedicated hardware such as one or more integrated circuits performing the functions or sub functions described. Such dedicated hardware may include graphic processors, digital signal processors, or one or more microprocessors and associated memories.

Industrial Applicability

It is apparent from the above that the embodiments of the invention are applicable to the broadcasting of multimedia content and descriptions and are of direct relevance to the computer, data processing and telecommunications industries.

The foregoing describes only some embodiments of the present invention, and modifications and/or changes can be made thereto without departing from the scope and spirit of the invention, the embodiments being illustrative and not restrictive.

Claims:

1. A method of forming a streamed presentation from at least one media object having content and description components, said method comprising the steps of:

generating a presentation description from at least one component description of said at least one media object; and

processing said presentation description to schedule delivery of component descriptions and content of said presentation to generate elementary data streams associated with said component descriptions and content.

2. A method according to claim 1 wherein said processing further comprises arranging said component descriptions into multiple ones of said data streams.

3. A method according to claim 1 wherein said presentation description comprises references to said description components and said description components are streamed with said at least one media object.

4. A method according to claim 1 wherein said presentation description is formed by importing said description components, and said generation operates to stream only said presentation description and said at least one media object.

5. A method of forming a streamed presentation of at least one media object having content and description components, said method comprising the steps of:

providing a presentation template that defines a structure of a presentation description;

applying said template to at least one description component of at least one associated media object to form said presentation description from each said description component; and

stream encoding said presentation description with each said associated media object to form said streamed presentation, whereby said at least one media object is reproducible using said presentation description.

6. A method of forming a presentation description for streaming content with description, said method comprising the steps of:

providing a presentation template that defines a structure of a presentation description;

applying said template to at least one description component of at least one associated media object to form said presentation description from each said description component, said presentation description defining a sequential relationship between description components desired for streamed reproduction and content components associated with said desired descriptions.

7. A method according to claim 6 further comprising applying said presentation description to the method of claim 1.

8. A method according to claim 1, 5 or 6 wherein said streamed presentation comprises a description tree having at least one node referencing a description object.

9. A method according to claim 8 wherein said streamed presentation further comprises at least one further node referencing at least one said media object.

10. A method according to claim 1, 5 or 6 wherein said stream encoding comprises:

parsing said presentation description to form a plurality of presentation sequential description objects, each said description object being associable with at least one associated media object; and

forming a streamed sequence of said description objects and related said associated media objects, said streamed sequence being said streamed presentation.

11. A method according to claim 10 wherein a relationship between said description objects and said associated media objects is defined by further objects forming part of said streamed presentation, each said further object comprising a tree structure having nodes each referencing at least one of said description objects and said media objects.

12. A method according to claim 1, 5 or 6 wherein said presentation description comprises an XML document describing content intended for reproduction in a time sequential manner.

13. A method according to claim 1, 5 or 6 wherein said presentation description is formed by modifying an SMIL description used to specify the timing and synchronization of said media objects and said descriptions

14. A streamed presentation comprising a plurality of content objects interspersed amongst a plurality of description objects, said description objects comprising references to multimedia content reproducible from said content objects.

15. A streamed multimedia presentation comprising a first stream representing a tree structure of said presentation, at least one second stream having object descriptors each referenced from said tree structure, at least one third stream comprising content referenced from said object descriptors and intended for reproduction in said presentation, and at least one fourth stream comprising descriptions of said content referenced from said object descriptors.

16. A streamed presentation according to claim 15 wherein said third stream comprises an MPEG-4 stream.

17. A streamed presentation according to claim 16 wherein said second stream comprises an Object Content Information stream having URI's referencing MPEG-7 information represented in said fourth stream.

18. A method of delivering an XML document, said method comprising the steps of:

dividing the document to separate XML structure from XML text; and

delivering said document in a plurality of data streams, at least one said stream comprising said XML structure and at least one other of said streams comprising said XML text.

19. A method according to claim 18 wherein said dividing comprises converting said XML documents into a tree representation.

20. A method according to claim 19 wherein said tree representation is divided in a breadth-first manner.

21. A method according to claim 19 wherein said tree representation is divided in a depth-first manner.

22. A method of processing a document described in a mark up language, said method comprising the steps of:

separating a structure and a text content of said document;

sending the structure before the text content; and

commencing to parse the received structure before the text content is received.

23. A method according to claim 22, further comprising the step of ignoring the received text content if it is found not to be required or unable to be interpreted as the result of parsing the corresponding structure.

24. A method according to claim 23, wherein said ignoring step comprises inhibiting a buffering of the text to be ignored.

25. A method according to claim 22, wherein the mark up language is XML.

26. A method according to claim 22, wherein said separating step comprises encoding the structure and the text content as two separate streams.

27. A method according to claim 26 wherein said document is formed as a tree hierarchy representation and said separating step further comprises interpreting said document in a depth-first fashion to form said two streams.

28. A method according to claim 26 wherein said document is formed as a tree hierarchy representation and said separating step further comprises interpreting said document in a breadth-first fashion to form said two streams.

29. Apparatus for performing the method of any one of claims 1 to 12 or 17 to 28.

30. A computer readable medium, having a program recorded thereon, where the program is configured to make a computer execute a procedure form a streamed presentation, said procedure being according to the method of any one of claims 1 to 12, or 17 to 28.

31. A method of forming a streamed presentation having streamed description substantially as described herein with reference to Figs, 2, 3, and 4C of the drawings.

32. A method of forming a streamed presentation having streamed description substantially as described herein with reference to Figs, 2, 5, and 4C of the drawings.

33. A streamed presentation substantially as described herein with reference to Fig. 4B or 4C of the drawings.

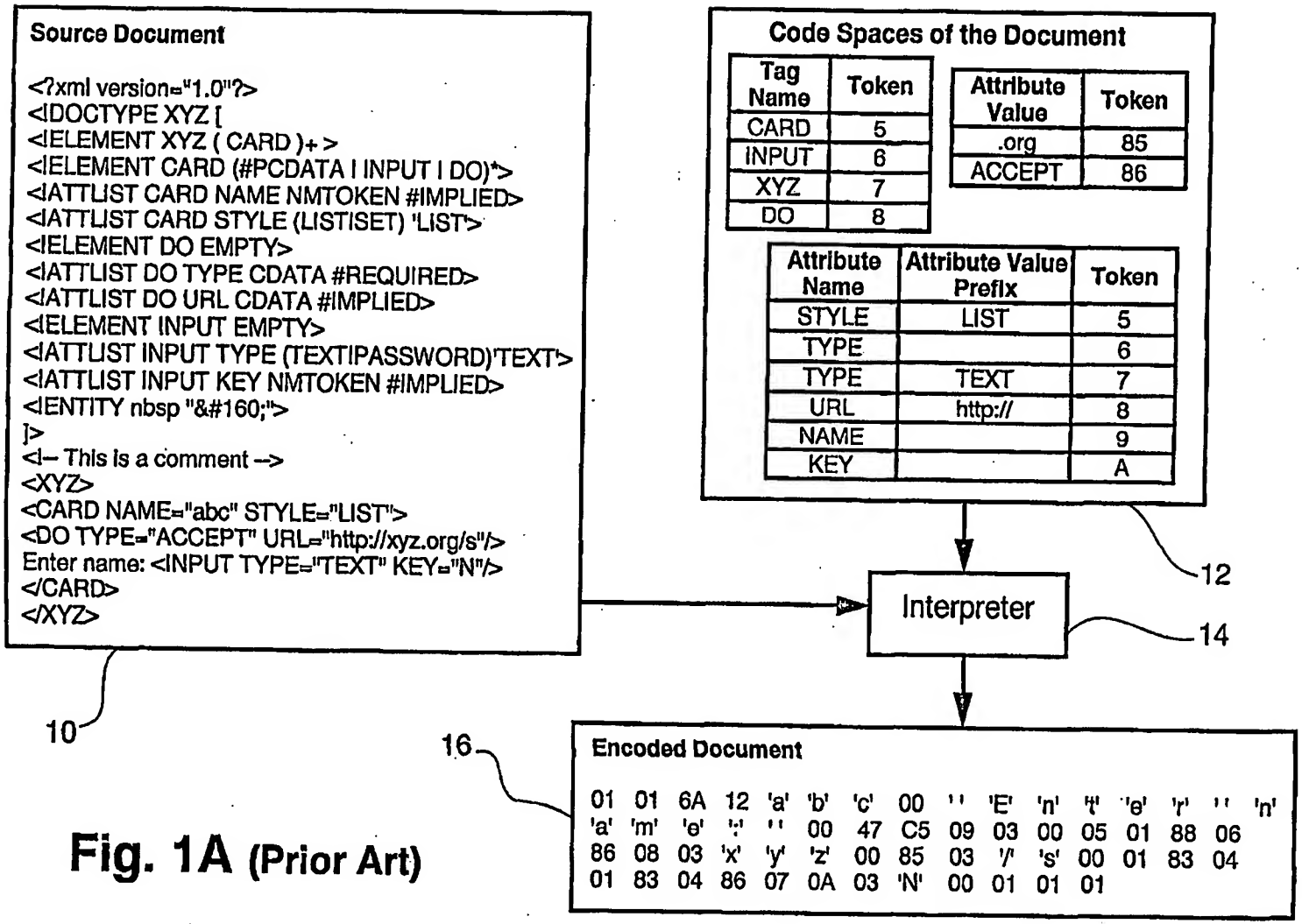


Fig. 1A (Prior Art)

Token in Stream	Description
01	Version number - WBXML version 1.1
01	Unknown public identifier
6A	charset=UTF-8 (MIBEnum is 106)
12	String table length
'a', 'b', 'c', 00,' ', 'E', 'n', 't', 'e', 'r', ' ', 'n', 'a', 'm', 'e', ':', ' ', 00	String table
47	XYZ, with content
C5	CARD, with content and attributes
09	NAME=
83	String table reference follows
00	String table index
05	STYLE="LIST"
01	END (of CARD attribute list)
88	DO, with attributes
06	TYPE=
86	ACCEPT
08	URL="http://"
03	Inline string follows
'x', 'y', 'z', 00	string
85	".org"
03	Inline string follows
'/', 's', 00	string
01	END (of DO attribute list)
83	String table reference follows
04	String table index
86	INPUT, with attributes
07	TYPE="TEXT"
0A	KEY=
03	Inline string follows
'N', 00	String
01	END (of INPUT attribute list)
01	END (of CARD element)
01	END (of XYZ element)

Fig. 1B (Prior Art)

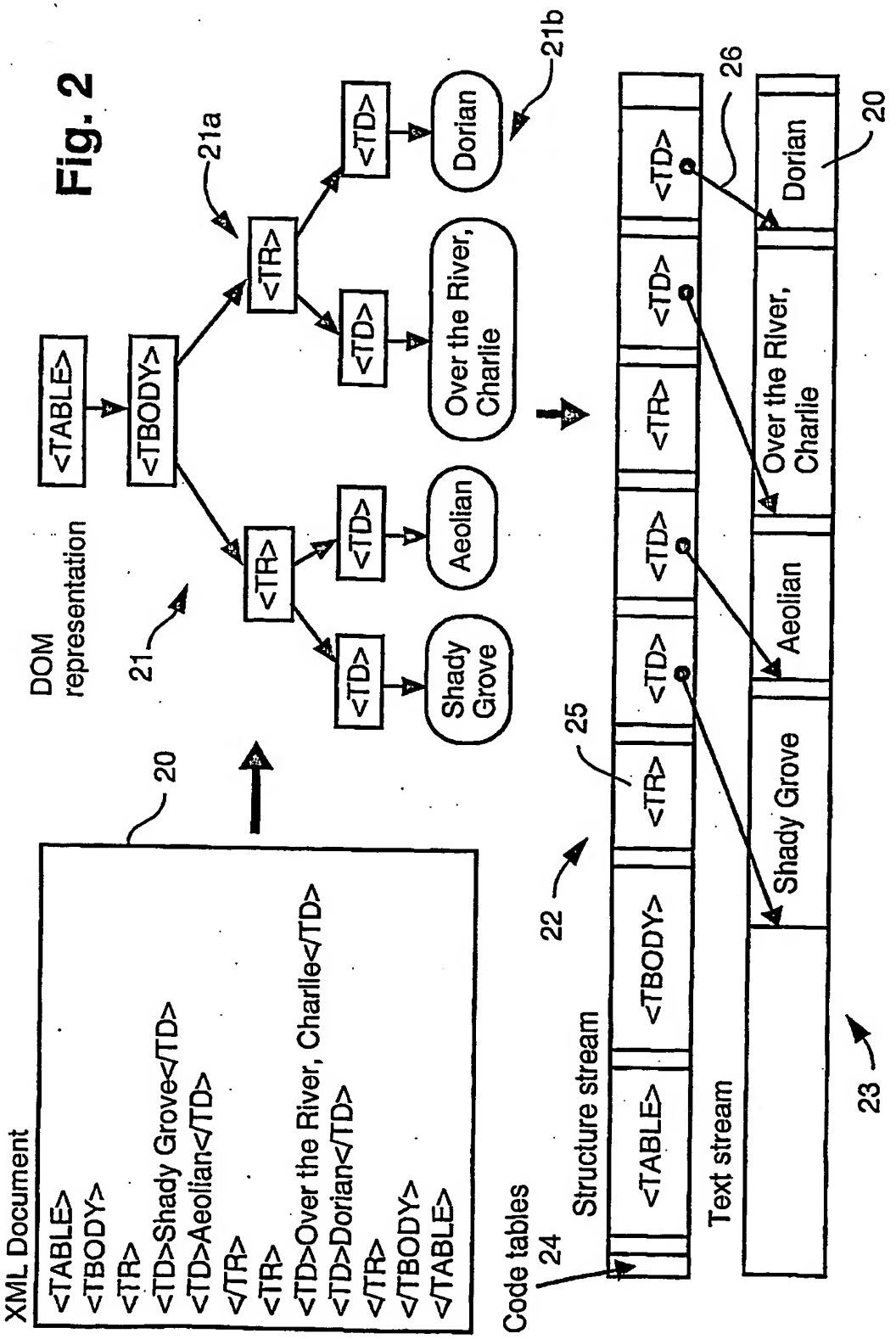

XML Document
<TABLE>
<TBODY>
<TR>
<TD>Shady Grove</TD>
<TD>Aeolian</TD>
</TR>
<TR>
<TD>Over the River, Charlie</TD>
<TD>Dorian</TD>
</TR>
</TBODY>
</TABLE>
20
DOM representation
21
<TABLE>
<TBODY>
<TR>
<TD>
Shady Grove
<TD>
Aeolian
<TR>
<TD>
Over the River, Charlie
<TD>
Dorian
21a
21b
Code tables
24
Structure stream
22
<TABLE>
<TBODY>
<TR>
<TD>
25
Text stream
23
Shady Grove
Aeolian
Over the River, Charlie
Dorian
26
20

Fig. 2

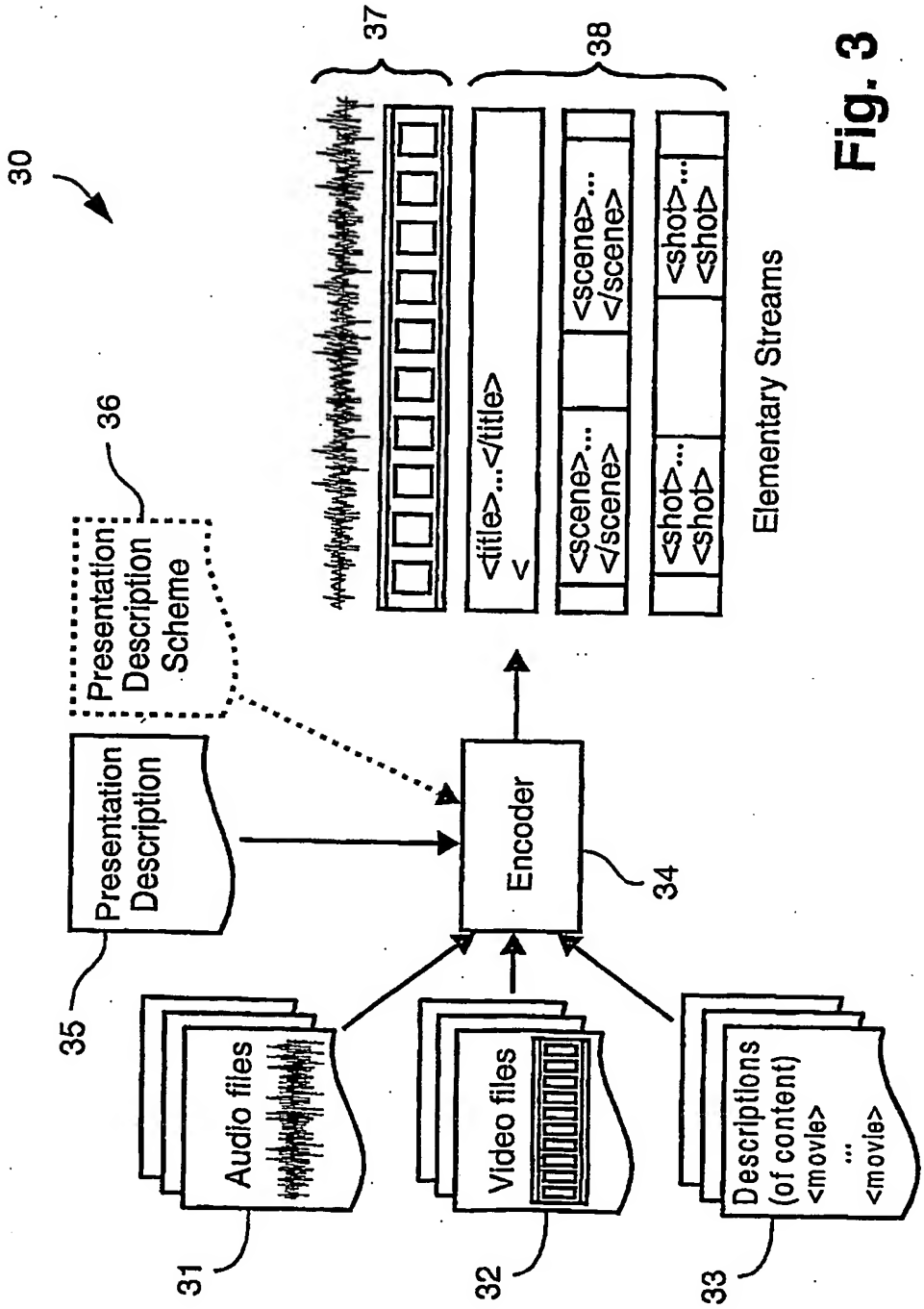

30
Presentation Description Scheme
36
Presentation Description
35
Audio files
31
Video files
32
Descriptions (of content)
<movie>
...
<movie>
33
Encoder
34
37
38
<title>...</title>
<scene>...
</scene>
<scene>...
</scene>
<shot>...
<shot>
<shot>...
<shot>
Elementary Streams

Fig. 3

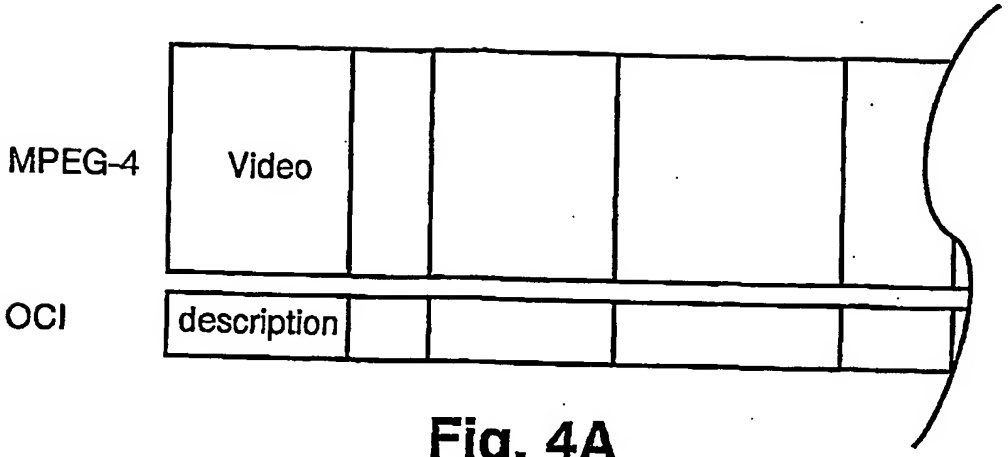


Fig. 4A
(Prior Art)

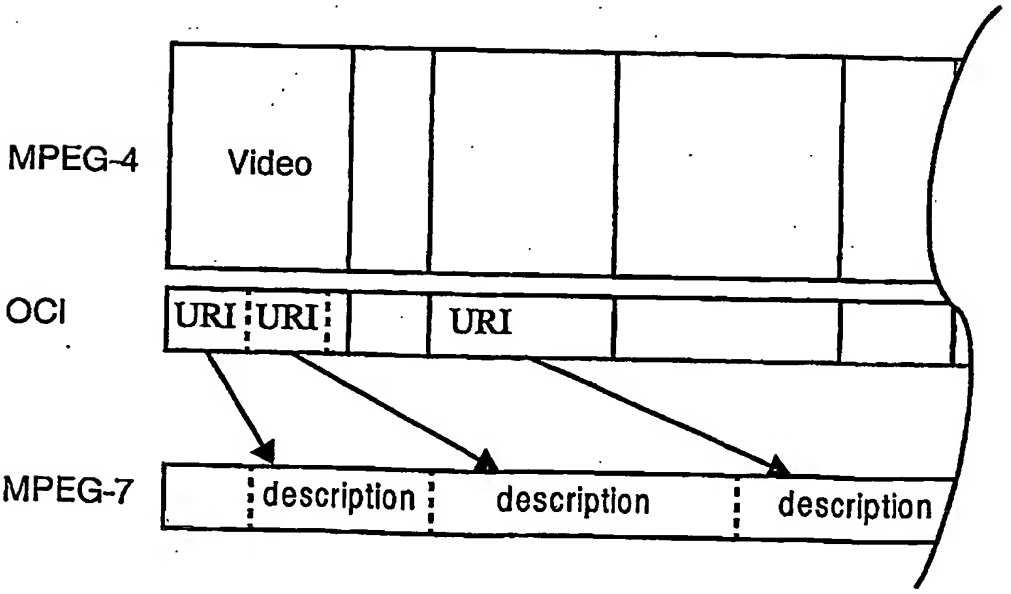


Fig. 4B

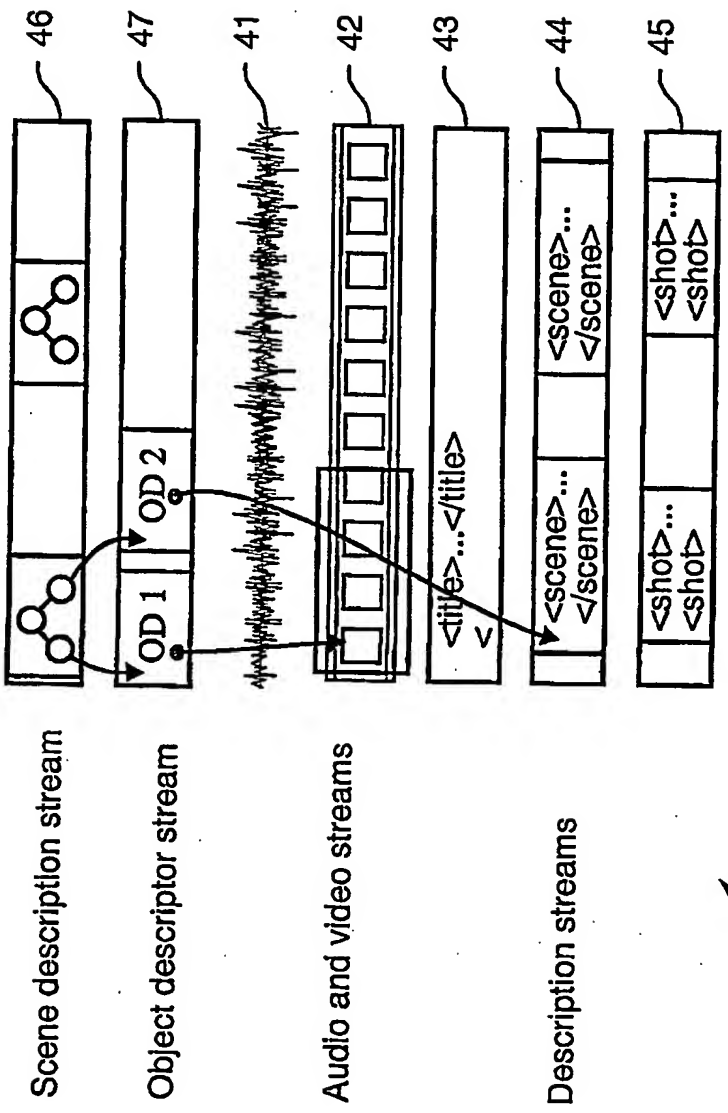

46
47
41
42
43
44
45
OD 1
OD 2
<title>...</title>
<scene>...
</scene>
<scene>...
</scene>
<shot>...
<shot>
<shot>...
<shot>
Scene description stream
Object descriptor stream
Audio and video streams
Description streams
Elementary Streams
40

Fig. 4C

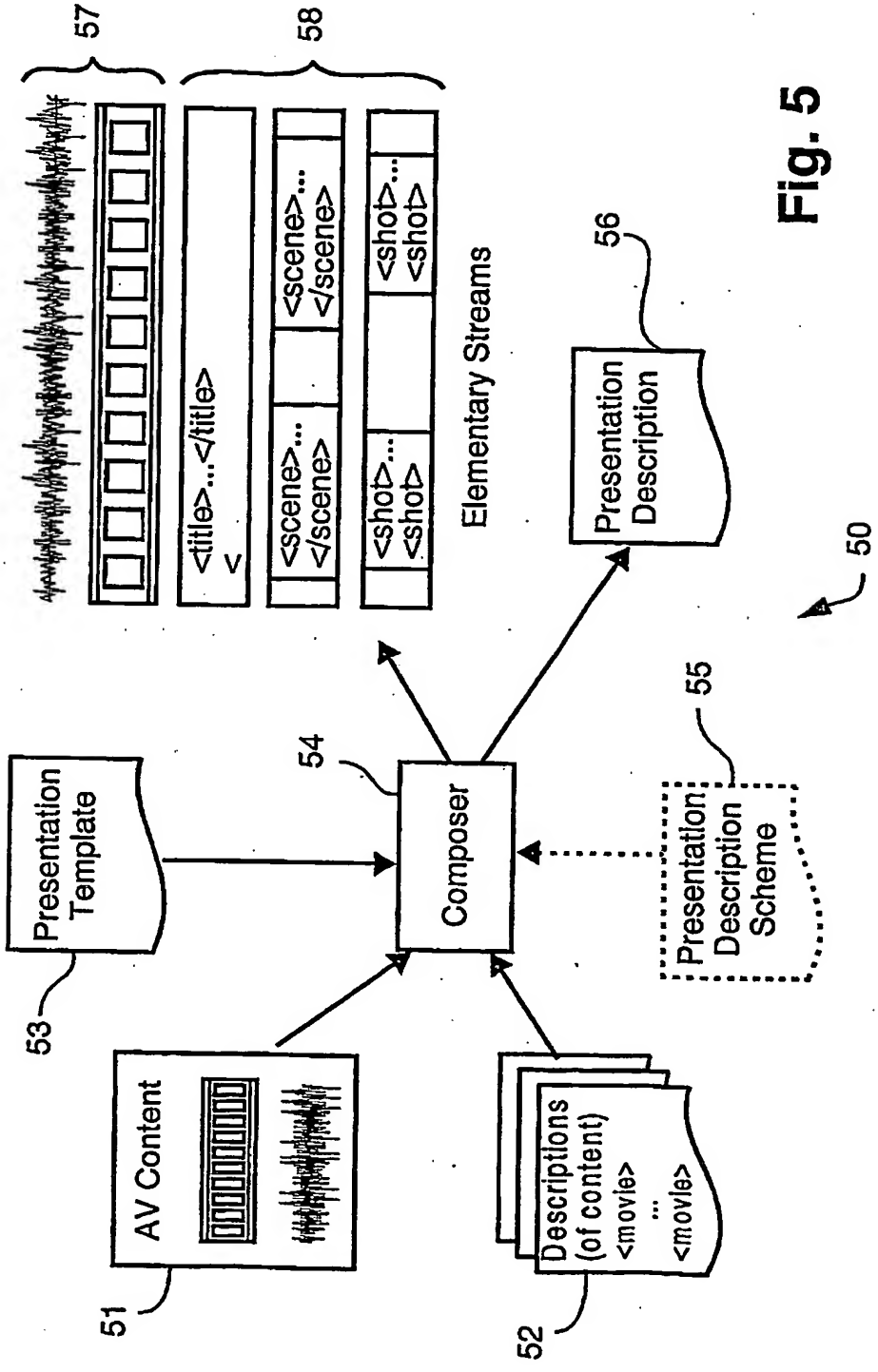

57
58
<title>...</title>
<scene>...
</scene>
<scene>...
</scene>
<shot>...
<shot>
<shot>...
<shot>
Elementary Streams
Presentation Description
56
Presentation Template
53
Composer
54
Presentation Description Scheme
55
AV Content
51
Descriptions (of content)
<movie>
...
<movie>
52
50

Fig. 5

Fig. 6(a)

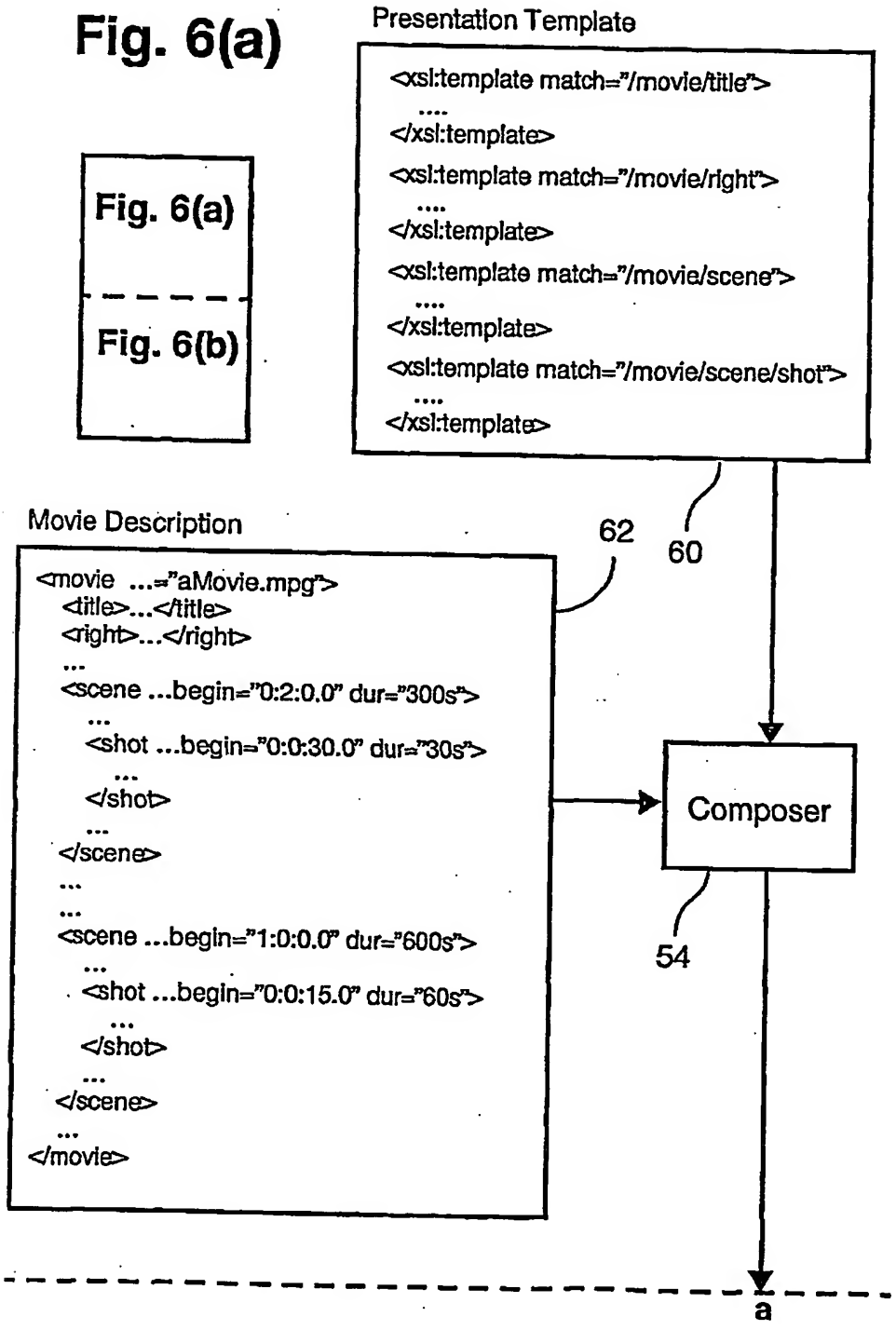


Fig. 6(a)
Fig. 6(b)
Presentation Template
<xsl:template match="/movie/title">
....
</xsl:template>
<xsl:template match="/movie/right">
....
</xsl:template>
<xsl:template match="/movie/scene">
....
</xsl:template>
<xsl:template match="/movie/scene/shot">
....
</xsl:template>
62
60
Movie Description
<movie ...="aMovie.mpg">
<title>...</title>
<right>...</right>
...
<scene ...begin="0:2:0.0" dur="300s">
...
<shot ...begin="0:0:30.0" dur="30s">
...
</shot>
...
</scene>
...
...
<scene ...begin="1:0:0.0" dur="600s">
...
<shot ...begin="0:0:15.0" dur="60s">
...
</shot>
...
</scene>
...
</movie>
Composer
54
a

a

Presentation Description

```
<par>
  <video src="aMovie.mpg" />
  <par begin="aMovie.begin"
       end="aMovie.end">
    <mpeg7 src="aMovieDesc.xml#xpointer(/movie/title)" content="/movie/title"/>
    <mpeg7 src="aMovieDesc.xml/#xpointer(/movie/right)" content="/movie/right"/>
    <par begin="2:0" dur="300s">
      <mpeg7 content="/movie/scene/">
        <!-- non shot-specific scene description -->
        ...
      </mpeg7>
      <par begin=0:30" dur="30s">
        <mpeg7 src="aMovieDesc.xml/#xpointer(/movie/scene[@begin='0:2:0.0']/shot[@begin="0:0:30:0])"
               content="/movie/scene/shot"/>
      </par>
    </par>
    ...
    <par begin="1:0:0" dur="600s">
      <mpeg7 content="/movie/scene/">
        <!-- non shot-specific scene description -->
        ...
      </mpeg7>
      <par begin="15:0" dur="660s">
        <mpeg7 src="aMovieDesc.xml/#xpointer(/movie/scene[@begin='1:0:0.0']/shot[@begin="0:0:15:0])"
               content="/movie/scene/shot"/>
      </par>
    </par>
    ...
  </par>
</par>
```

64

Fig. 6(a)

Fig. 6(b)

Fig. 6(b)

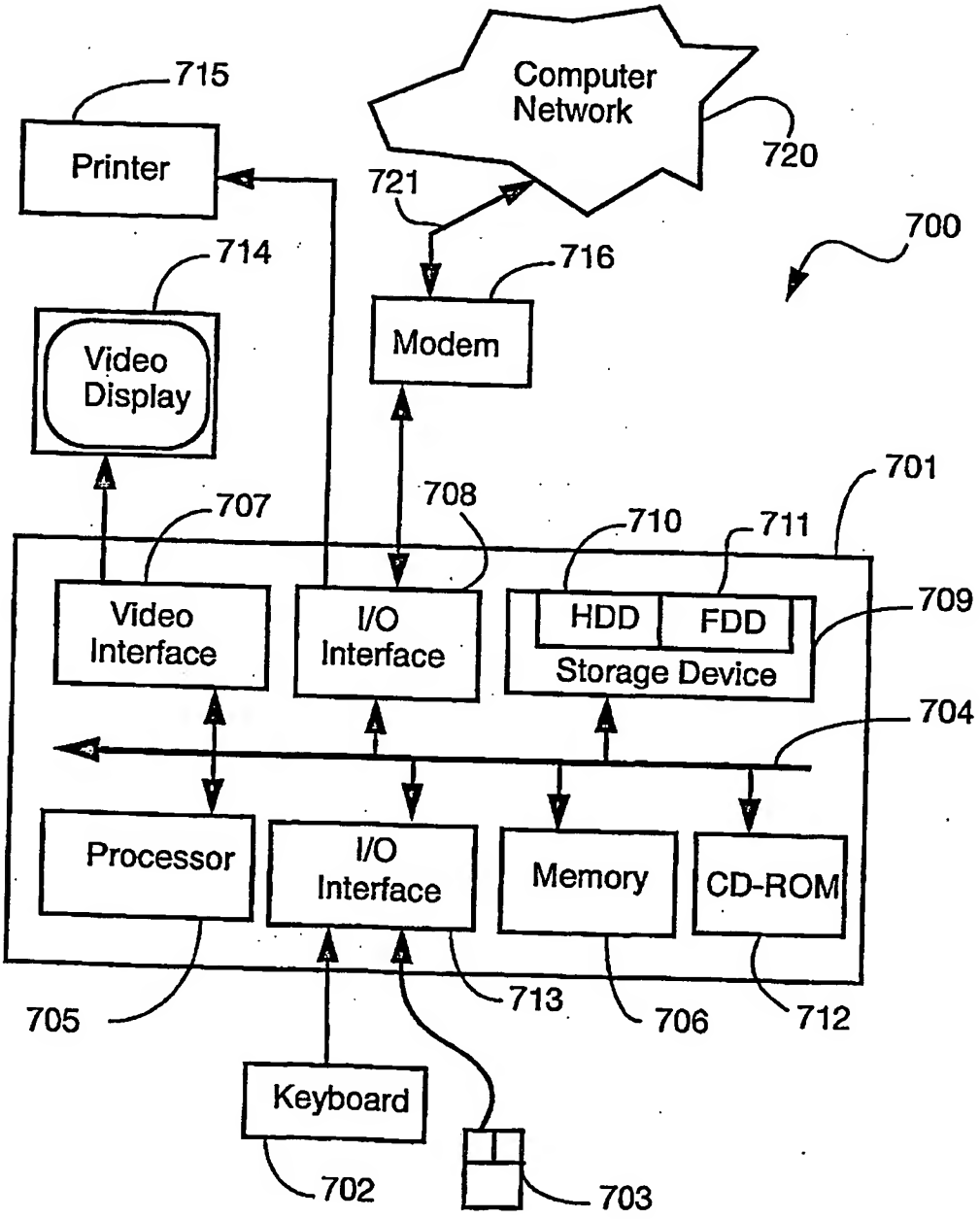

Computer Network
720
715
Printer
721
700
714
716
Video Display
Modem
701
707
708
710
711
Video Interface
I/O Interface
HDD
FDD
709
Storage Device
704
Processor
I/O Interface
Memory
CD-ROM
705
713
706
712
Keyboard
702
703

Fig. 7

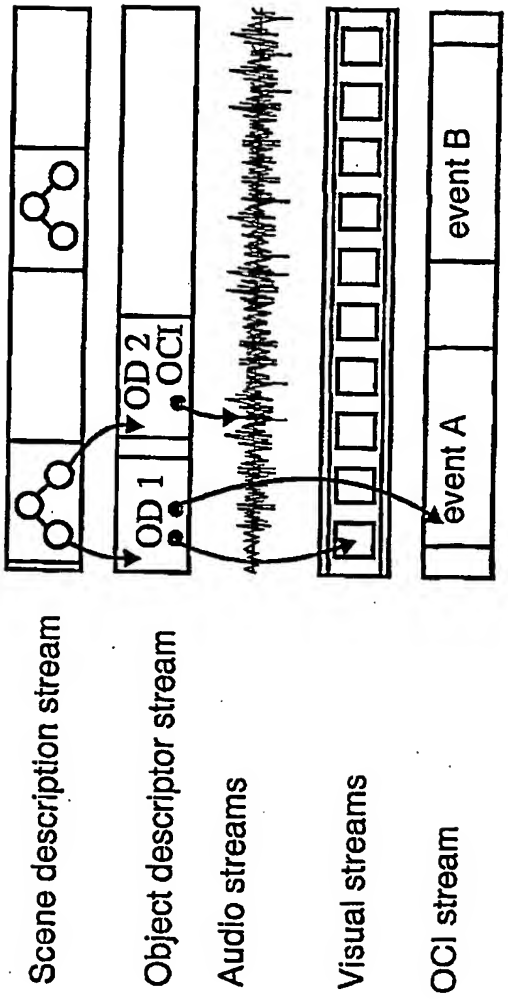

Scene description stream
Object descriptor stream
Audio streams
Visual streams
OCI stream
OD 1
OD 2
OCI
event A
event B
Elementary Streams

Fig. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/00799

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: G06F 9/45, 15/16, 17/00, 17/30, 17/60

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F, H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

AU: IPC AS ABOVE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPAT, INSPEC, USPTO: XML, TEXT, STRUCTURE, CONTENT, MULTIMEDIA, SEPARATE, STREAM PRESENTATION.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X P, Y	*EFFICIENT REPRESENTATION AND STREAMING OF XML CONTENT OVER THE INTERNET MEDIUM, By GIRARDOT AND SUNDARESAN, 2000 IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO. ICME2000 PROCEEDINGS. LATEST ADVANCES IN THE FAST CHANGING WORLD OF MULTIMEDIA. Pt. Vol. 1, pp.67-70. New York, July/August 2000.*	1, 5, 6, 14-18, 22 2-4, 7-13, 19-21, 22-30
X, Y	*XML: NOT A SILVER BULLET, BUT A GREAT PIPE WRENCH, By USDIN AND GRAHAM, MULBERRY TECHNOLOGIES INC., ROCKVILLE, MD STANDARDVIEW, vol. 6, no. 3 pp. 125-132, September 1998.*	1-30
X, Y	US 6083276 A (DAVIDSON et al.) 4 July 2000 Entire document, see: abstract, columns 1-4, drawings	1-30

[X] Further documents are listed in the continuation of Box C [X] See patent family annex

* Special categories of cited documents:

"A" Document defining the general state of the art which is not considered to be of particular relevance

"E" Earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
5 September 2001

Date of mailing of the international search report
10 SEPTEMBER 2001

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaustralia.gov.au
Facsimile No. (02) 6285 3929

Authorized officer

CHARLES BERKO
Telephone No : (02) 6283 2169

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/00799

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	US 6012098 A (BAYEH et al.) 4 January 2000	
X	Entire document	1-5, 14, 15, 18, 22
Y	Entire document,	11, 19, 27, 28
Y	abstract, column 4, lines 1-22	23
	CA 2255047 A (IBM CANADA LTD) 30 May 2000	
X	Entire document	1-6, 14, 15, 18, 22
Y	Entire document	7-13, 16, 17, 19-21, 23-30
	AU 53031/98 A (DUDLEY) 27 August 1998	
X	Abstract, page 2	14
Y	Entire document	1-13, 15-30
	US 6025876 A (BRAMLEY) 15 February 2000	
Y	Entire document	16, 17
	WO 2000/24195 A (MORECOM INC.) 27 April 2000	
Y	Entire document, see abstract, pages 1-3.	16, 17
	WO 2000/48088 A (ALYSIS TECHNOLOGIES INC.) 17 August 2000	
P, Y	Abstract, page 3	23
Y	*GRAPHIC ELEMENT MARKUP, By ROBERTS M. XML INFOSPHERE INC. FORT WORTH, TX, USA, XML EUROPE '99 CONFERENCE PROCEEDINGS, pp. 547-570, GRANADA, SPAIN, 26-30 APRIL 1999, SPONSORED BY OASIS, W3C WORLD WIDE WEB CONSORTIUM.*	1-30

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/00799

Box I **Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos :

because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos : 31-33

because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

They do not comply with rule 6.2(a) as they make references and rely on the description and drawings

3. ☐ Claims Nos :

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II **Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

SEE EXTRA SHEET (Supplemental Box)

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU01/00799

Supplemental Box
(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: II

1. Claim 1 directed to forming streamed presentation, is characterised by delivery to generate content associated data streams.

2. Claim 5, also, directed to forming streamed presentation, is characterised by a structure defining template and stream encoding in association with a reproducible media object.

3. Claim 6 directed to forming presentation description for streaming content, is characterised by a structure defining template applied to description components and associated content components sequential relationships.

4. Claim 14 directed to streamed presentation, ***is extremely broadly and speculatively*** characterised by description objects that reference reproducible multimedia content interspersed with content objects; - ***as is the case with all Multimedia Presentation using any Markup Language.***

5. Claim 15 directed to streamed multimedia presentation is characterised by at least four tree structure representing streams each having object descriptor referenced content.

6. Claim 18 directed to XML document delivery, is characterised by document division into separate XML structure and XML text and data stream delivery of document with at least one stream comprising the structure and at least one other comprising the text.

The commonality between these claims lies in **streamed document presentation**. This feature is not novel. As demonstrated by the cited documents, **streamed document presentation** is very well known in use of Markup languages. These claims therefore lack unity *'a posteriori'*.

7. Claim 22 directed to directed to processing a document described in Markup language, is characterised by structure and text content separation and parsing of received structure before text content is received.

   While this claim has **Markup language and structure and text content separation in common** with only claim 1, it does not have any feature(s) **in common** with the other independent claims.

Thus the feature of **streamed document presentation** is considered to be "a first technical" feature, while **Markup language, and structure and text content separation** is considered to be "a second technical" feature.

Since the groups of claims identified do not share either of the technical features identified, no "technical relationship" exists between them, as required by PCT rule 13.2. Accordingly, the international application does not relate to one invention or to a single inventive concept.

However, the PCT rule notwithstanding, any tree structure involves some form of branching/branches one way or the other and, as it were, involves ***separation of file content*** that would be considered to be inherently ***streamed***. This being the case, claim 22 could be said to have ***streamed document presentation*** **in common** with all the other independent claims. This feature has been deemed to be not novel. Thus the claims as a whole would further lack unity *'a posteriori'*.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU01/00799

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
US	6083276	NONE					
US	6012098	NONE					
CA	2255047	NONE					
AU	53031/98	NONE					
US	6025876	EP	747856	FR	2735258	JP	9102952
WO	200024195	AU	200012107	EP	1121802		
WO	200048088	AU	200033634				

END OF ANNEX